The Need (or not) for Server-side Operation Standards

Steve Hankin, NOAA/PMEL

GO-ESSP

Oct. 2009, Hamburg, Germany

Today (mostly)

"move the data to the analysis"

e.g. ftp files or download subsets to your local host for analysis

As data volumes and complexity grow (notably CMIP5)

~~"move the data to the analysis"~~

"move the analysis to the data"

reduce network transfers, exploit local data performance, parallel processing, specialized operations, automation of complex data management, …

In ESG

We have 2 classes of server-side operation:

1. Axis-collapsing operations
    - definite integral, average, extrema, variance, …)
2. Regridding operations
    - Including GRIDSPEC to lat/long regridding

(Requirements for server-side computation of diagnostic quantities on native grids - tbd)

We could imagine many others:

- 1st&2nd derivatives, vorticity, divergence, EOFs, FFTs, …

Classes of server-side operations
(based upon degree of client control)

1. No control -- 100% pre-configured
   - e.g. model diagnostic zonalMomentumAdvection = U*dU/dx
2. Control the endpoints
   - e.g. low/high limits of integration
3. Control the expression
   - e.g. “(U-Ubar)^2 + (V-Vbar)^2”
4. Specify algorithmic code
   - e.g. upload scripts

(ESG offers 2-3 using a variant of OPeNDAP syntax.)

1. No control -- 100% pre-configured
   - e.g. “diff” computed as “airt-sst”
2. Control the endpoints
   - e.g. low/high limits of integration
3. Control the expression
   - e.g. “(U-Ubar)^2 + (V-Vbar)^2”
4. Specify algorithmic code
   - e.g. upload scripts

So far attempts to start a conversation on a standardized syntax haven’t gotten legs.

Questions:

1. Does the community have requirements for server-side operation standards?

2. If so what are they? How do we advance them?